

# Fake News Detection

(<https://youtu.be/voFIKCTHCRA>)

Anushya Subbiah  
[anushya@stanford.edu](mailto:anushya@stanford.edu)

Divya Sudhakar  
[divyasud@stanford.edu](mailto:divyasud@stanford.edu)

Kenny Hsu  
[shsu1@stanford.edu](mailto:shsu1@stanford.edu)

## Introduction

We'd like to reliably and scalably detect fake news - articles masquerading as news containing intentional misinformation - using the linguistic features in the news as well as user engagement features from Twitter.

We experimented with a variety of deep neural networks such as MLPs, RNNs, transfer learning using BERT.

These models were able to achieve higher accuracy than a human. The models are able to achieve high enough accuracy with just the news features. User engagement features improve accuracy but only marginally.

## Data

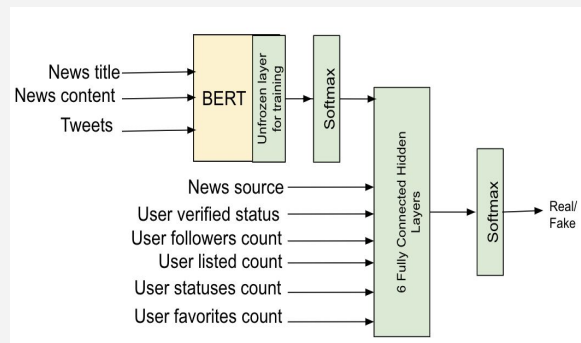
[FakeNewsNet](#): Labeled News datasets from PolitiFact and GossipCop joined with data on user engagement with the news articles from Twitter.

- 22,233 articles total.
- Approximately 75% real news, 25% fake.
- 1.8 million tweets along with user data.

## Features

- Title and text of news content
- Source (domain name) of news article
- Text of the tweet
- Screen name, id, location of user who tweeted
- Social data on the user (# of followers/friends, how many tweets, etc...)

## Model: BERT + DNN



## Results

- Accuracy: 84%
- F1 score:0.89

## Future Work

- Incorporating images from news article as features.
- Incorporating retweets and spread/virality of the news on twitter as features.
- Other approaches to aggregating the tweets associated with a news article

## Discussion

Human baseline: Manually labeling approximately 360 examples in the test set achieved an accuracy of 79.1% and recall of 36%.

RNN using only the title features was able to fit the training set fairly well. In general, just the news features is sufficient for good accuracy. Accuracy gain from user engagement features was marginal.

User engagement data was valuable in accurately classifying some examples. Model struggled to correctly classify articles from wikipedia without the user data. Model still struggles with celebrity gossip - something which even humans struggle with without additional research.

## References

- [1] [Fake News Detection on Social Media: A Data Mining Perspective](#)
- [2] [FakeNewsNet: A Data Repository with News Content, Social Context and Spatiotemporal Information for Studying Fake News on Social Media](#)
- [3] [Fake News vs Satire: A Dataset and Analysis](#)
- [4] [A Retrospective Analysis of the Fake News Challenge Stance Detection Task](#)
- [5] [“Liar, Liar Pants on Fire”: A New Benchmark Dataset for Fake News Detection](#)
- [6] [BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding](#)